The Use of Surrounding Visual Context in Handheld AR: Device vs. User Perspective Rendering

Klen Čopič Pucihar
SCC, Lancaster University, UK
k.copicpucihar@lancaster.ac.uk

Paul Coulton
LICA, Lancaster University, UK
p.coulton@lancaster.ac.uk

Jason Alexander
SCC, Lancaster University, UK
alexander@lancaster.ac.uk

ABSTRACT
The magic lens paradigm, a commonly used descriptor for handheld Augmented Reality (AR), presents the user with dual views: the augmented view (magic lens) that appears on the device, and the real view of the surroundings (what the user can see around the perimeter of the device). The augmented view is typically implemented by rendering the video captured by the rear-facing camera directly onto the device's screen. This results in dual perspectives—the real world being captured from the device's perspective rather than the user's perspective (what an observer would see looking through a transparent glass pane). These differences manifest themselves in misaligned and/or incorrectly scaled transparency resulting in the dual-view problem.

This paper presents two user studies comparing (a) device-perspective and (b) fixed Point-of-View (POV) user-perspective magic lenses to analyze the effect of the dual-view problem on the use of the surrounding visual context. The results confirm that the dual-view problem, a result of dual perspective, has a significant effect on the use of information from the surrounding visual context. The study also highlights that magnification and not the dual-view problem is the key factor explaining the correlation between magic lens size and the increased intensity of the magic lens type effect. From the results, we derive design guidelines for future magic lenses.

Author Keywords
AR; magic lens; dual views; dual-view; user-perspective;

ACM Classification Keywords
H.5.1. Information interfaces and presentation (e.g., HCI): Artificial, augmented, and virtual realities.

INTRODUCTION
Handheld AR on mobile devices is usually implemented using the magic lens paradigm [6], where the device acts as a transparent glass pane showing a digitally enhanced view of the scene laying behind the pane. A fundamental characteristic of such AR implementation is that there are in fact two views available to the user, the augmented view (magic lens) and the real view (what user can see around the perimeter of the device). In ideal transparency, the magic lens view exactly matches with its surroundings thus the two views seamlessly merge. However, handheld devices are not transparent, so virtual transparency is created by rendering the video stream from a single back-facing camera onto the device's forward-facing 2D display. Unfortunately, the imagery captured and rendered by the device does not match the ideal transparency—what would be seen when looking through a clear glass pane.

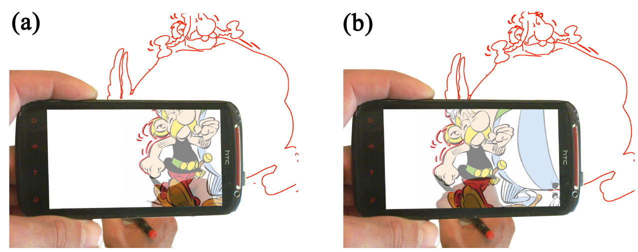


Figure 1: The dual-view problem illustrated on AR-supported sketching example: (a) Device- (b) User-perspective view.

The dual-view problem occurs when ideal transparency is not achieved. Figure 1a illustrates device-perspective rendering (note how the magic lens imagery does not match with the surroundings) and Figure 1b illustrates user-perspective rendering (note the exact match between rendering and the surroundings). The perspective differences between the device and user are the main cause of transparency distortion and manifest in incorrect scaling and misalignment (Figure 1a). This may impede the user's ability to: (1) Merge the magic lens view with the wider surroundings (*view-merge interaction*); and (2) Efficiently complete tasks that require crossing border between the lens and the surroundings (*cross-context interaction*).

View-merge interaction is important in scenarios with high information density such as AR on printed media (e.g. navigation and route planning on augmented paper maps [25], digitally enhanced text documents [18], and augmented magazine games such as crossword puzzle hints). Cross-context interaction is important when the user's hand interacts within the AR workspace and crosses the boundaries between real and augmented views. Examples include AR to support user sketching (see Figure 1), transcribing digital instructions/routes to paper maps, or where tangible objects are brought into and out of the augmented view, e.g. AR chess with physical pieces.

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada

ACM 978-1-4503-2473-1/14/04…$15.00.
http://dx.doi.org/10.1145/2556288.2557125

Seamless view-merge and cross-context interactions are expected to facilitate more fluid interaction as less time will be wasted interpreting different views, allow easier use of potentially valuable surrounding information, provide greater understanding of information context, and aid the construction and retention of a mental model of the information space [4, 23].

This paper investigates the effect of the dual-view problem on surrounding context use by comparing two types of magic lens renderings: device-perspective rendering and fixed-POV user-perspective rendering. Additionally, we examine the role of physical magic lens size on the severity of the dual-view problem. We evaluate these factors using off-the-shelf mobile devices (i.e. phones and tablets) and focus on the class of tabletop-sized AR systems. These setups are expected to benefit most from user-perspective rendering because other transparency distortions, such as diplopia and limited depth-of-field, are minimized at small magic lens distances (denoted as d' on Figure 2a).

This paper contributes: (1) A thorough analysis of the dual-view problem, identifying diplopia (double vision) as a further issue with potential to effect the use of the surrounding context; (2) Two user studies providing empirical evidence that the dual-view problem significantly affects the use of the surrounding context; (3) Analysis identifying that magnification (not the dual-view problem) explains the correlation between the magic lens size and the increased intensity of the magic lens type effect; (4) Design guidelines to support optimal magic lens selection for future AR design.

PROBLEM ANALYSIS

The dual-view situation, a common characteristic of handheld AR setups, may result in the dual-view problem if the magic lens transparency does not match that of a clear glass pane. In handheld AR, the magic lens transparency is typically implemented by rendering the video stream captured by a single back-facing camera onto the device's 2D screen. This is known as device-perspective rendering and does not match the ideal transparency suggested by the magic lens paradigm. The difference between the ideal and this commonly implemented virtual transparency arise from: (1) Monocular camera scene capture and rendering, and distorting binocular disparity; (2) Rendering the captured scene on the screen which is closer to the observer than the captured environment causing the depth-of-field problem; and (3) The camera capturing the real-world scene from a different perspective than the observer's creating a dual-perspective view. This section explores each of these factors in detail.

Distortion of Binocular Disparity

The observer's view, commonly referred to as the direct view, is stereoscopic, whereby the left and right eye capture two horizontally offset images. The difference in perceived location of an object caused by this offset is known as binocular disparity.

Single camera (monocular) scene capture and rendering distorts binocular disparity causing depth-perception [8] and diplopia problems. Contrary to diplopia (discussed in the following section), depth-perception is not expected to have an effect on the use of the surrounding visual context. Additionally, more important depth cues such as depth ordering and motion parallax [1, 14] can be used to correctly interpret scene depth.

The Diplopia or Double vision Problem

When the binocular disparity of two horizontally offset images exceeds a certain threshold, diplopia occurs (see Figure 2d). The point at which this occurs is known as the diplopia threshold [11]. To avoid diplopia, eyes converge [10] and meet at a single point in space.

Diplopia commonly occurs in handheld AR systems as distorted binocular disparity produces different disparity planes for the magic lens and the world as shown in Figure 2e. The different disparity planes mean the user's eyes can converge at either the magic lens (foreground) or the real-world object (background), but not on both. This increases binocular disparity of the background (on Figure 2e colored pink) or foreground object (on Figure 2e colored black). Such an effect increases with magic lens distance (d' on Figure 2e) and when it exceeds the diplopia threshold, double vision of the one view occurs. Thus diplopia results in diminishing image clarity of the one view, potentially affecting the ability to merge the magic lens view with its surroundings (see Figure 2d).

Restricted Depth-of-Field of the Human Eye

Handheld AR systems are also affected by the restricted depth-of-field of the human eye, which is defined as the distance between the nearest and furthest objects in a scene that appear acceptably sharp. The typical implementation of virtual transparency has the effect of transposing the world scene closer to the observer. The limitations of the depth-of-field of the human eye makes it difficult to keep the magic lens and the surrounding context in perfect focus resulting in blurring of the surrounding context or the magic lens. This problem becomes more pronounced as the distance between the magic lens and the scene grows larger, potentially contributing to the difficulty of relating the magic lens view to the surrounding context. However, the restricted depth-of-field can also be beneficial as it makes diplopia-affected areas of the observer's FOV less distracting [7].

The Dual-Perspective Situation

The dual-perspective situation is a result of typical handheld AR setups which is caused by perspective differences between the camera and the observer such as: (1) *Non-centered Scene Capture:* while the rear-facing camera is typically positioned off-center from the phone's screen, the video stream is still presented in the center of the display. This can be observed in Figure 2a, where the blue dot in the scene is rendered at the position of the red dot; (2) *Differences in the Field of View (FOV)*: the phone

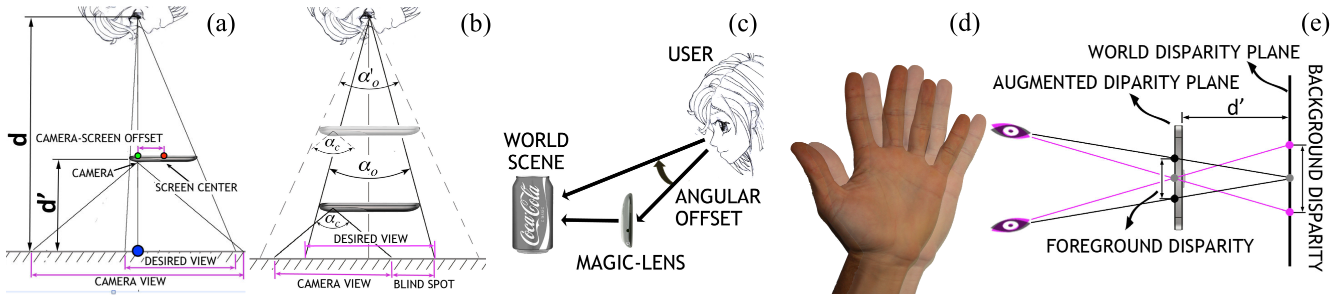


Figure 2: (a) Non-center Scene Capture: Device-perspective view: Blue dot is rendered to the position of red dot. User-perspective view: Blue dot is rendered at the expected position (green dot). (b) Camera FOV is static (α_c), user-perspective FOV is dynamic and changes with magic lens distance (from α_o to α'_o). (c) Angular offset. (d) Double vision/diplopia reducing image clarity. (e) Converging eyes at foreground (magic lens, pink) or background (world, black) introducing binocular disparity.

camera's FOV (α_c on Figure 2b) is different to the FOV the observer would see if the magic lens acted as a transparent glass pane (α_o in Figure 2b). In contrast to the dynamic FOV of the transparent glass pane, the camera has a static FOV (in Figure 2b, the transparent glass pane FOV changes with lens distance from α_o to α'_o); and (3) *Angular Offset of Views:* when the device is held at a non-perpendicular angle to the observer's POV as shown in Figure 2c.

These factors cumulatively contribute to misalignment and/or incorrect scaling of AR imagery as shown in Figure 1a. While the dual-perspective problem can be solved by user-perspective rendering (see Figure 1b), device-perspective rendering remains the current industry standard.

Problem Summary

The dual-view problem is a cumulative result of: the dual-perspective situation, a restricted depth-of-field, and the diplopia problem. This paper focuses on the dual-view problem, a consequence of dual-perspective as: (1) Previous work has shown that the dual-view problem, caused by dual-perspective situation, has a significant effect on users' spatial perception [3, 9] and for this reason we expect it to also have a strong influence on the use of the surrounding visual context; (2) The severity of all but the dual-perspective problem is reduced by decreasing the magic lens distance (d' on Figure 2a); and (3) Misalignment and incorrect scaling are the most obvious distortions causing the dual-view problem.

RELATED WORK

Issues associated with the dual-view problem such as different disparity planes, the dual-perspective view, and limited depth-of-field are well known in the AR community [10, 16]. However, there are few user studies examining the effects of these issues on the usability of handheld AR interfaces. Understanding the effects of the dual-view problem is vital to enhance the usability of handheld AR interfaces which Olsson et al. identified as inconsistent and questionable for pragmatic usefulness in everyday life [22].

User-Perspective Rendering

User-perspective magic lens implementations are significantly more complex to implement than device-perspective implementations, however they solve the dual-perspective problem. While device-perspective rendering only requires the relative position of the device camera in relation to the environment (camera pose registration [15, 17]), user-perspective rendering also requires: (1) the relative position of the observer's head in relation to the magic lens (head pose tracking [2]); and (2) an up-to-date high-density 3D model of the environment.

Scene Reconstruction

In vision-based AR systems, 3D models of the environment are commonly available as they are required for camera pose registration [15, 17]. These models are typically predefined or built incrementally through multi-frame tracking, meaning they are unable to cope with fast changing environments. To date, there are no solutions that allow accurate real-time 3D mapping of dynamic environments on a handheld device, limiting user-perspective rendering solutions to static environments.

User-perspective Rendering Prototypes

Baričevič et al. implemented a geometrically correct user-perspective rendering prototype on a purpose built hardware platform [3]. They performed per-pixel 3D scene reconstruction using a Kinect depth camera sensor. This provided a real-time per-pixel 3D scene, however, limits in the Kinect's depth sensor range prevent correct rendering of the user's hand while interacting with the AR scene. Until depth sensors become available on mobile devices such a method is unsuitable for handheld deployment.

Recently, several simplified variations of user-perspective rendering have emerged on handheld devices [9, 13, 19]. Hill et al. [13] and Matsuda et al. [19] implemented user-perspective rendering by trimming the back-facing camera's video stream based on the viewer's POV defined by using the front-facing camera. This allowed correct alignment to the viewer's POV at a single distance, a consequence of using camera-captured images taken from device's POV, creating imagery with different perspective distortion to those seen by the observer. The considerable computational overhead of head pose tracking and that only a single video stream is available on the majority of handheld devices means such solutions have yet to be realized on a single mobile device.

To date, Čopič Pucihar et al.'s fixed-POV user-perspective rendering is the only implementation that runs on off-the-shelf hardware. They reduced the problem complexity by eliminating head pose tracking by constraining the observer's POV above the center of the device screen. With the user's cooperation, in terms of maintaining their head position relative to the device, the system creates user-perspective imagery with correct perspective distortion [9].

Surrounding Context

During handheld AR interaction, the magic lens obscures only a fragment of the observer's FOV leaving a large segment to view the surrounding context of the real world. This surrounding context holds valuable cues, and enables the user to retain a mental model of the information space, crucial in tasks such as large document navigation [4, 23].

Relating augmented content to the real world is not a trivial task for the user. Rohs et al.'s map navigation task which compared a magic lens [6] and a peephole [20] interaction on a mobile phone did not detect significant performance difference between these two interactions [25]. This suggests that users' tend to ignore the surrounding context (printed map) when interacting with device-perspective magic lenses. However, later studies examining the impact of item density [24] did find a significant effect from visual context. In both studies the dual-perspective situation was not considered as a potential factor affecting the user's spatial perception and the use of the surrounding context.

Outside AR, the utilization of surrounding context has primarily been explored through evaluation of focus+context visualizations such as fisheye views [12] or focus+context screens [5]. Such visualizations show a portion of the document in high resolution (the focus) while the surrounding document (the context) is rendered with lower resolution. In contrast to the more common overview+detail, focus+context visualizations allow seamless merging of the two views resulting in faster and more accurate extraction of information from a large document that does not fit on screen [4]. We consider this context as analogous to AR's surrounding context and focus analogous to the magic lens. Therefore, by facilitating easier merging of the two views in handheld AR we are likely to produce a similar effect.

The Dual-Perspective Situation

The main cause of the dual-view problem is the dual-perspective situation which was found to significantly distort users' spatial perception [3, 9]. Baričević et al. [3] explored the effect of the dual-view problem on users' spatial perception by comparing user and device-perspective rendering in search and selection tasks in an MR simulator. In the selection task they found a significant affect of magic lens type only on tablet-sized magic lenses. Building on Baričević et al.'s work, Čopič Pucihar et al. [9] conducted real-world user studies to explore users' initial expectations and ability to deal with distorted spatial perception by comparing device-perspective and a fixed-POV user-perspective magic lens in a selection task. Their results confirmed a significant effect of dual perspective on users' spatial perception. For this reason we expect it to also influence the use of the surrounding visual context.

FIXED-POV USER-PERSPECTIVE RENDERING

Fixed-POV user-perspective rendering is a technique that produces user-perspective imagery without the need for head pose tracking. This is done by fixing the observer's POV above the center of the magic lens. This does not prevent the observer from moving the phone in z-direction or from viewing the scene at an angle as long as the user follows the instructions shown in Figure 3e. The fixed-POV technique was used because at the time of running the study full user-perspective rendering was not yet feasible on a single handheld device. With recent reports of successful adoption of such a view [9], where a detailed description can be found, we believe this is currently the best method for evaluating the dual-view problem.

Implementation Limitations

Fixed-POV user-perspective rendering constrains users to holding the phone perpendicular to their POV. As a result, angular offset does not contribute to the dual-perspective problem. In a tabletop-sized environment, the distance between the magic lens and the interactive workspace (*d'* on Figure 3e) is expected to remain small, reducing the potential effect of angular offset. We also assume holding the phone perpendicularly is the most intuitive interaction.

Fixing the observer's POV in relation to the device screen also reduces the motion parallax effect. However, acquiring different views of the scene, through collaborative movement of the observer with the device, is still possible and so motion parallax may still be used as a depth cue. As depth perception is not important when trying to merge the surrounding context with the magic lens, this limitation is not deemed significant within the context of this study.

Finally, the system uses predefined textures or textures generated incrementally on the fly as models of the planar 3D scene. Irrespective of the method used, such maps are not updated with every captured frame. Real-time per-pixel 3D reconstruction is not yet realizable on handheld devices which results in occlusion of any introduced objects not part of the mapped 3D scene. Although solutions that trim the camera's captured images can visualize dynamic scenes, the correct alignment of the users POV is only possible at a single distance, therefore, dynamic scene elements will end up misaligned and incorrectly scaled. Until alternative scene reconstruction methods appear, or depth cameras are introduced onto mobile devices, geometrically correct user-perspective rendering remains limited to static environments, which is sufficient for our evaluations of the use of surrounding context.

EXPERIMENTAL APPROACH

Due to complexity of evaluating surrounding context use, we specifically chose to conduct two tightly constrained experiments. In study A, cross-context interaction was evaluated on a drag-and-drop task forcing users to cross the lens boundary, analogous to AR examples where user's hand interacts within AR workspace and crosses context border (for examples see introduction). In study B, view-merge interaction was evaluated in a map navigation task where path following required decision making. Even though the designed task is artificial, this interaction pattern is analogous to navigation and route-planning on augmented paper maps. Additionally, as Oh and Hua used a similar methodology, it was deemed appropriate [21].

USER STUDY A: CROSS-CONTEXT INTERACTION

The study is designed to investigate how the dual-view problem, a result of the dual-perspective situation, affects cross-context interaction. We also investigate how magic lens size affects the dual-view problem by comparing cross-context interaction tasks using a device-perspective magic lens and a fixed-POV user-perspective magic lens. Regardless of magic lens size, we expect participants to: (H1) Perform better with the fixed-POV user-perspective lens; and (H2) Prefer the fixed-POV user-perspective lens.

Experimental Design

We asked users to perform a series of drag-and-drop actions where objects are brought across the contextual border from the surrounding context into the magic lens. This direction of contextual border crossing was chosen to allow us to capture users' performance immediately post border crossing, thus allowing us to evaluate any effects of the crossing.

The experiment is a 2×2 factor within-subjects design with the independent variables being *size* and *type* of the magic lens. The magic lens size had two values: (1) *Phone*—4.3 inch display (HTC Sensation); and (2) *Tablet*—10.1 inch display (Motorola XOOM). The magic lens type, took two values: (1) *Device-perspective magic lens*—showing the view captured from the handheld device's camera; and (2) *Fixed-POV user-perspective magic lens*—described in the previous section.

Experimental Apparatus and Setup

The physical setup (see Figure 3a) consisted of a 32-inch plasma screen laid horizontally at table height (note: in the remainder of the paper we refer to this setup as a *tabletop surface*). The tabletop surface is a replacement for printed media which allows us to render different static images and the virtual representation of the hand. The virtual hand is controlled via a mouse on a surface immediately to the right of the plasma screen and adjusted to each participant's elbow height. Throughout the experiment, participants stand and hold the magic lens in their left hand.

Replacing The User's Hand with a Virtual Hand

The mouse controlled virtual hand is used for two reasons: (1) As noted in the related work section, geometrically correct user-perspective rendering of dynamic scenes is currently not possible on handheld devices; and (2) It removes the users' direct spatial perception of their hand, meaning they rely only on visual feedback, which is expected to highlight the effects of the dual-view problem.

Magic lens Orientation

The physical placement of the rear-facing camera differs between the two devices. To match the direction of misalignment caused by the camera-screen offset we asked participants to hold the phone-sized magic lens in portrait orientation and the tablet-sized magic lens in landscape orientation. This keeps the direction of misalignment the same between devices thus isolating any differences of the varying form factor between the two lenses. The horizontal camera offset of both devices is 1 cm, whereas the vertical offset of the tablet is 7.9 cm and the phone 4.3 cm.

Experimental Task

Each task requires the participant to drag a ball across the context border into an augmented basket using the mouse controlled virtual hand as shown in Figure 3c. The ball and basket are placed according to the following conditions: (1) The ball is always generated 30–40 cm from the basket and outside the initial magic lens view; and (2) The basket is always generated in the right half of the tabletop surface.

Participants were instructed to complete the task as quickly as possible, with a minimum accuracy of 0.5cm (defined through pilot tests where participants demonstrated such level of performance is easily achievable). The ball changed to a green color when this accuracy was achieved. Performance was measured by task completion time and path deviation. The time measurement begins as the ball is picked up and stops when it is dropped into the basket. The approach path of the drag-and-drop action is recorded at a frequency of 15Hz from which path deviation is calculated as a proportion between the length of the approach path and the length of a direct line linking initial ball position and basket.

Experimental Procedure

Participants first completed a setup activity in which the distance between the observer's POV and the interactive surface was measured (*d* on Figure 3e) to allow us to initialize the fixed-POV user-perspective rendering. Afterwards, they were introduced to handheld AR and fixed-POV rendering by demonstration and written instructions (see Figure 3e). Participants always started with the small magic lens and performed seven repetitions with both magic lens types. A questionnaire was completed before moving to the tablet-sized lens. The order of lens type was counterbalanced.

Participants

The study was completed by 15 participants (4 female, 11 male); all were right handed and aged between 24 and 45.

Results

In total, 420 task repetitions were recorded (15 participants $\times$ 2 magic lens sizes $\times$2 magic lens types $\times$7 repetitions).

Time and Path Deviation

Statistical analysis only partly confirmed H1 of improved performance with the user-perspective magic lens.

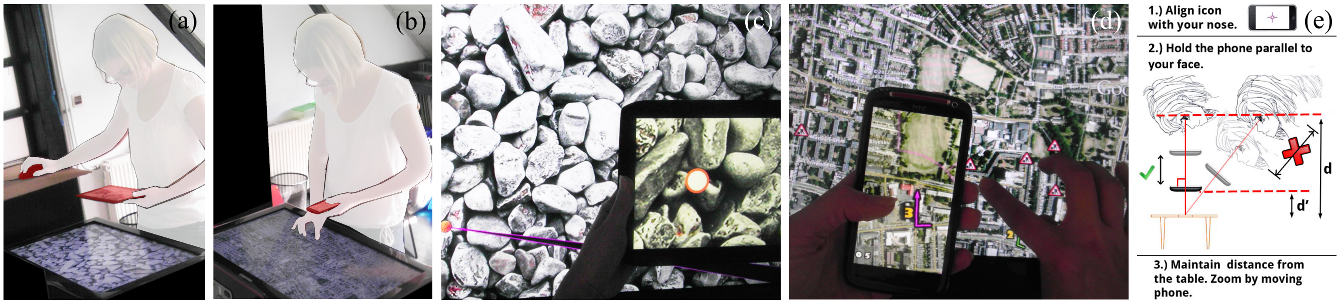


Figure 3: (a) Study A: experimental setup. (b) Study B: experimental setup. (c) Selection task of Study A. The orange ball on the left is brought into the basket visible on the magic lens view. (d) Path following task of Study B. Counting corners by tapping on the phone screen. (e) Instructions page summarizing fixed-POV user-perspective rendering assumptions.

A repeated measures ANOVA with two factors (magic lens type and size) showed a significant effect of magic lens type on users' performance (see Figure 4a and 4b). The fixed-POV user-perspective magic lens shows a significant reduction in task completion time ($F_{1,14} = 21.229$, $p < 0.001$) and path deviation ($F_{1,14} = 23.063$, $p < 0.001$) over the device-perspective lens.

There was no significant effect of magic lens size on task completion time, however, there was a significant effect on path deviation ($F_{1,14} = 19.812$, $p = 0.001$) with the large lens exhibiting greater deviation.

An interaction between magic lens type and size was detected in both task completion time ($F_{1,14} = 15.512$, $p = 0.001$) and path deviation ($F_{1,14} = 23.811$, $p = 0.001$).

Post-hoc analysis showed that magic lens type only has a significant effect for tablet-sized displays. Participants performed better in completion time ($p < 0.001$, 95% CI [0.33, 0.90] seconds) and path deviation ($p < 0.001$, 95% CI [0.12, 0.36]) when using fixed-POV user-perspective rendering on tablet-sized lenses. Such effect was not detected in case of the phone-sized lens.

User Preferences

A Wilcoxon signed rank test partially confirmed H2, that the fixed-POV user-perspective magic lens would be preferred as shown in Figure 4c. Significant preference was only detected for the tablet-sized magic lens ($p = 0.001$). Of the 13 participants expressing a preference for fixed-POV user-perspective rendering: 6-participants ~67% (95% CI [51%, 82%]) related this to smaller magnification; 2-participants ~22% (95% CI [8%, 36%]) stated an advantage of having no blind spots—areas of the scene that are occluded by the magic lens but not rendered; and 1-participant ~11% (95% CI [0.6%, 21%]) related it to better spatial perception.

USER STUDY B: MERGING VIEWS

The second of the two studies investigated how the dual-view problem, a result of dual-perspective situation, affects participants' ability to merge information on the magic lens with the surrounding context (view-merge interaction). Our predictions are identical to user study A.

Experimental Design

To explore users' ability to merge the magic lens view with the surrounding visual context participants were asked to perform a path following task in which they counted the number of corners from start to finish. We used an identical experimental design as User Study A (with the exception of only testing phone-sized magic lenses).

Experimental Setup

The experimental setup is similar to that used in User Study A except we remove the virtual hand to prevent its use to establish a link between the magic lens view and the surrounding context. If participants are unable to effectively use the surrounding context without the hand to establish a reference point, then this indicates the inability to easily merge the two views.

Experimental Task

Each task required participants to navigate a path on a map while counting the number of corners from start to finish. Each path has only one possible solution and is defined by: a starting point, three instruction arrows, roadwork signs, and a finishing point. The first and third instruction arrows (colored in pink) were presented on the lens, while the start, finish, and the second instruction arrow (colored in green) were presented on the tabletop surface. Roadwork signs are used to block other path solutions (see Figure 5a).

Participants count the number of corners by tapping on the phone screen. After each tap a sound is played to acknowledge the counter increase displayed in the bottom corner of the magic lens as shown Figure 3d. The task is completed by tapping on the count number, at which point the solution is shown. During the task, three time measurements were performed: (1) *Task-time*—the time from when the task is shown on the tabletop surface until the last corner is counted; (2) *Startup-time*—the time from when the task is shown until the instruction arrow 2 is reached; and (3) *Finishing-time*—the time required to navigate from second to third instruction arrow.

Task startup-time is of interest as it was expected to be the most difficult part of the task because participants see the map for the first time and might experience difficulties orientating themselves. The finishing-time is similar to the startup-time as both task segments require the user to merge

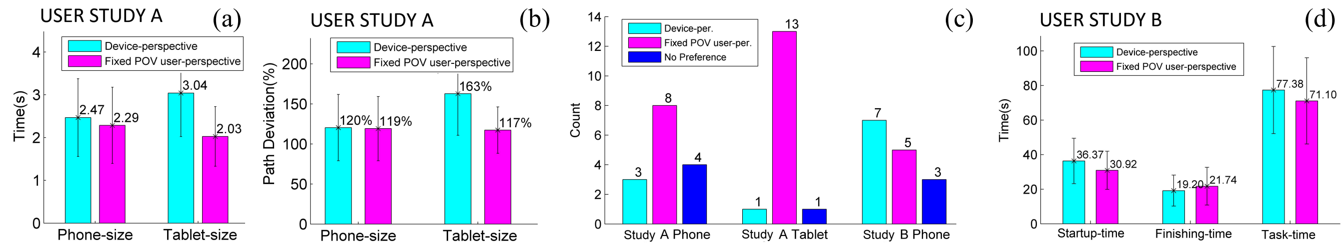


Figure 4: (a) Study A: Selection task time completion. (b) Study A: Selection task path deviation. (c) Magic lens preference choice for User Study A and B. (d) Study B: Path following: *startup-time*, *finishing-time* and *task-time* for a phone-sized lens.

the magic lens instructions with the surrounding map. However, towards the end of the task it was expected users to have become more familiar with the map.

We used multiple maps and paths to ensure map learning did not take place. To achieve comparable difficulty, all paths: (1) Have the same number of instruction arrows; (2) Flow from the right to the left side of the tabletop surface; (3) Have equivalent path lengths (±10%); (4) Have from 11 to 14 cornering points; and (5) Are generated on maps with similar information density.

Participants and Experimental Procedure

All participants who completed Study A also completed Study B. Before starting Study B, the differences between the two types of magic lens were specifically highlighted using a map texture. Participants then experimented with both magic lens types for a maximum of 2 minutes. Afterwards, participants were given the task description and demonstration after which they performed a trial run with each magic lens type. They then moved to the timed tasks where they completed six paths, three in each of the two magic lens types. Path assignment to magic lens type was counterbalanced. After completing all repetitions they were asked to name their primary task completion strategies, select their preferred lens type and justify their choice.

Results

All 90 task repetitions were successfully completed (15 participants × 2 magic lens types × 3 repetitions).

Task, Startup, and Finishing Time

Paired sample T-tests partially confirmed H1, that fixed-POV user-perspective rendering was faster as shown in Figure 4d. The average task-time was not significantly quicker with fixed-POV user-perspective rendering. However, when broken down into its component parts, the startup-time for fixed-POV user-perspective rendering was significantly faster ($p = 0.029$, 95% CI [0.65, 10.24] sec). However, once participants were more familiar with the map, the finishing-time showed no significant difference.

Preference

A Wilcoxon rank test showed no significant preference for either rendering although two additional people preferred device- over user-perspective rendering (see Figure 4c).

Observational Results

Through participant observation and their feedback we identified three main strategies when performing merge-view interaction, namely: (1) *Image comparison*—landmarks from the magic lens render were matched to the same landmarks in the real-world scene; (2) *Spatial awareness combined with image comparison*—spatial awareness was used to narrow down search regions for matching landmarks (i.e. landmarks on the phone should lay below the phone); (3) *Simultaneous view*—edges of the magic lens were merged with the surrounding view. For device-perspective rendering, the dominant strategy for all participants was image comparison. For fixed-POV user-perspective rendering, participants showed diversity in the strategies employed with 8-participants using image comparison ~53% (95% CI [66.2%, 40.5%]), 3-participants using spatial awareness combined with image comparison ~20% (95% CI [9.6%, 30.3%]), and 4-participants using simultaneous view ~27% (95% CI [15.2%, 38.1%]) strategy.

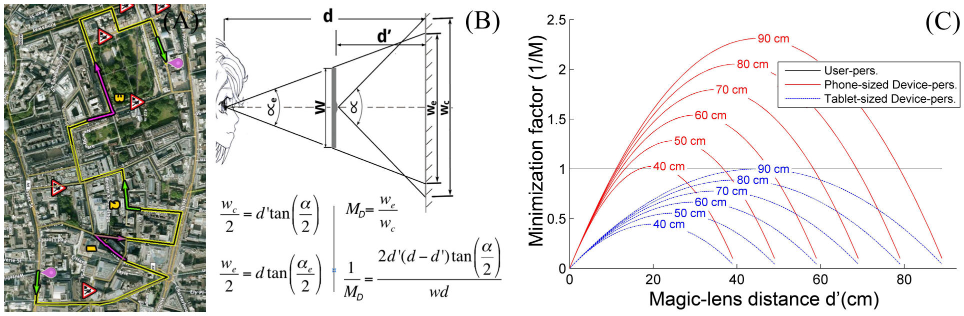
We observed that when participants only used the image comparison strategy they wanted to move away from the surface to get a wider view of the scene. For strategies where spatial perception was used participants wanted to move closer to the surface, especially in the case of simultaneous viewing. Unfortunately, due to frustration caused by poor tracking at close proximity many participants abandoned the simultaneous view strategy.

USER-CENTRIC AR DESIGN GUIDELINES

Results of user Study A showed: (1) a significant effect of lens type; (2) a strong connection between lens type and size; and (3) a limited effect of lens size. The connection detected in (2) suggests that lens type and size are inter-related. Here we try to identify the cause of this connection and we start by analyzing the two main reasons participants listed when asked to justify their magic lens preference choice, namely: (a) magnification; and (b) presence of blind spots. In the following sub-sections we consider why these are important and how lens size and type interact to produce these undesirable effects.

Understanding Magnification and Blind spots

Magnifying magic lenses occlude a larger portion of the observer's FOV than that rendered on the screen, creating information blind spots—real-world areas that are occluded by the physical magic lens, but not visible on the magic-lens render as illustrated in Figure 2b. In addition to blind spots, lenses with larger magnification show a smaller fragment of the real-world scene thus reducing the number



(D)

LENS SIZE	ADVANTAGE	MAGIC-LENS TYPE	
		Device-pers.	**User-pers.**
Phone-size	Magnification Adv. (1/M>1)	**YES**	**NO**
	Spatial Adv. (no dual-veiw p.)	NO	YES
Tablet-size	Magnification Adv. (1/M>1)	**NO**	**YES**
	Spatial Adv. (no dual-view p.)	NO	YES

Figure 5: (A) Study B: task solution. (B) Derivation of device-perspective magnification (M_D) based on magic lens size (w), camera FOV (∝), the magic lens (d') and observer distance (d). (C) Minimization (1/Magnification) plot for device-perspective (blue and red) and user-perspective (black) magic lenses. (D) Guideline table for tabletop-sized setting for 4.3 and 10.1 inch magic lenses.

of available landmarks. Blind spots and reduction of landmarks are expected to undermine users ability to cross the contextual border and perform image comparison which was identified as the predominant strategy when merging views (see observational results of Study B).

Therefore, in cross-context and view-merge interaction we assume magnification is a disadvantage, so in the remainder of the paper we use the term *magnification advantage* to describe lenses that minimise—show a larger portion of the scene than is being occluded by the lens (see Figure 5b).

Magnification Model

The magnification model (Figure 5b) is derived from a theoretical analysis of: (1) Ideal transparency—what the observer would see when looking through a transparent glass pane (w_e on Figure 5b); and (2) Camera-generated virtual transparency (w_c on Figure 5b). Throughout this analysis we define magnification as a relation between ideal and camera-generated transparency ($M=w_e/w_c$).

To achieve user-perspective rendering, the camera-generated transparency is adjusted to match ideal transparency. This results in realistic rendering with no magnification ($w_e=w_c$, $M_U=w_e/w_c=1$). For device-perspective rendering, the camera-generated transparency is not adjusted. This results in varying magnification ($M_D=w_e/w_c$). Based on Figure 5b we can derive a magnification model for such device-perspective lens: $M_D=wd/(2d(d-d')\tan(\propto/2))$.

Utilizing the theoretical model, and the screen size (w), FOV of the camera (∝), distance to workspace (d'), and a variety of observer distances (d), we generated Figure 5c. For numerical reasons, Figure 5c shows minimization (the inverse of magnification, 1/M) and is used to illustrate when magnification and minimization occurs for the two devices used in our studies (minimising above 1, magnifying below 1).

By interpreting Figure 5c, it is possible to generate first and third row of the guideline table in Figure 5d. These two rows summarise when the magnification advantage occurs. For example: on Figure 5c the plot for phone-sized magic lens with device-perspective rendering (red plot) stays above 1 for most observer distances (d=40—90 cm), so the first cell in the guideline table indicates that for such a lens, device-perspective rendering has a magnification advantage over user-perspective rendering. The opposite is true in case of tablet-sized lenses (blue plot). Therefore, within the context of the experimental setup magnification advantage is dependent on magic lens type and size. This allows us to conclude that magnification is responsible for the interaction between device size and type detected in Study A.

Guidelines for Magic lens Selection

Our results show that the surrounding context use is affected by: (1) existence of the dual-view problem; and (2) magnification advantage. Based on these two parameters we propose a guideline table (see Figure 5d) which enables selection of the optimal magic lens type and can be generalized to any tabletop-sized AR workspace. To do this, the specific setup parameters are plugged into the model (Figure 5b), in order to produce the equivalent of Figure 5c. The magnification advantage (row 1 and 3, Figure 5d) is then derived, while the second criteria (row 2 and 4) is only lens type dependent.

Once the table is generated, it may be interpreted by using the following two rules: (1) The most appropriate rendering type is the one where magnification and no dual-view problem coexist; (2) When these advantages do not coexist, the designer needs to decide which property is more important. We only found conclusive evidence that the existence of the dual-view problem is more important than magnification in view-merge intensive interactions (see Study B results), so in such scenarios the user-perspective lens should be chosen. If user-perspective rendering is not an option the application design should aim to operate in observer and magic lens ranges where magic lens minimises (1/M>1).

DISCUSSION

The results of our studies show that the dual-view problem plays a significant role in the use of surrounding context.

Magnification: The link between Lens Type and Size

The quantitative results of cross-context interaction, User Study A, and the user study conducted by Baričević et al. [3] detected a connection between magic lens type and size.

The magnification model analysis uncovered magnification as the cause of this connection, coupled with participants' dislike for magnification and blind spots, this allows us to conclude: The strong significant effect of magic lens type in large lenses, detected in User Study A, is not a result of increased dual-view effect (a consequence of magic lens size), but rather the result of a cumulative effect of the dual-view problem and magnification (see table on Figure 5d). Consequently, applying a fisheye lens to the tablet-sized device-perspective magic lens would eliminate the magnification, reducing the effect of magic lens type.

Irrespective of the fact that in phone-sized user-perspective magic lenses, the magnification advantage is not present, fixed-POV user-perspective lenses continue to outperform device-perspective lenses in view-merge interactions (User Study B). However, evidence to support this was only found in the startup-time segment of the path following task. Here, orientation and decision making is expected to be most difficult because at the start of the task participants see the map for the first time and must orient themselves to begin decision making.

Inability to Merge Views

Fixed-POV user-perspective rendering brings virtual transparency closer to ideal transparency by solving the dual-perspective problem. However, the majority of participants continued to use image comparison as part of their main view merging strategy. This implies the magic lens was not used as part of the scene where the edges of the lens seamlessly merge with the surroundings. Therefore, for confident decision making, participants used additional visual verification.

This outcome is linked to the participants' lack of confidence in virtual transparency, a consequence of: (1) The accuracy of the provided imagery which is mainly affected by the participants' ability to align their POV to the one assumed by the system; and (2) Limited camera pose tracking capabilities at a close distance, preventing participants from moving the magic lens very close to the AR workspace (distances<12cm). At such distances other factors causing the dual-view problem, such as depth-of-field and diplopia do not exist.

Study Limitations

Testing Only Phone-sized Lenses

User Study B only tested phone-sized lenses. Our guidelines show that device-perspective tablet-sized magic lenses do not have any advantages over user-perspective lenses. There is a strong reason to assume that for tablet-sized magic lenses our predictions (H1 and H2) will be confirmed for view-merge interaction, justifying our decision to run Study B only on phone-sized lenses.

Restricting Users' Movement

The study limits magic lens interaction to relatively close proximity to the AR workspace. This limitation is required because: (1) Isolation of the dual-perspective situation is only possible by operating the magic lens in relatively close proximity; and (2) Aligning the observer's POV to the one anticipated by fixed-POV magic lens becomes more difficult at larger distances. The maximal distance was defined by a pilot study and identified at approximately 20cm, however, it was set to 25cm to enable comfortable operation around the detected distance.

We believe that (2) puts fixed-POV user-perspective magic lens at a disadvantage compared to the ideal dynamic-POV implementation. Hence, the observed advantage of user-perspective over device-perspective lens is bound to exist in ideal dynamic POV implementations.

Experimental Setup

Participants always stood during interaction, resulting in AR at close to top-down viewing. This helped participants to align their POV to the one anticipated by the user-perspective magic lens. However, the generalization of results is not limited to top-down cases as full user-perspective rendering solutions render geometrically correct user-perspective imagery without the need for alignment. The study is also limited to planar workspaces and 2D augmentations. This is reasonable and does not limit the applicability of our results because workspaces with high information density are commonly planar (i.e. maps, text documents). Also, as the study interaction mainly happened at close to top-down viewing, rendering 3D content would create little difference compared to 2D content.

Finally, the interaction space of the experiment setup has a restricted FOV. While this is problematic when considering outdoor settings, it does not limit the applicability of our results within the class of tabletop-sized AR workspaces where similar restrictions exist.

Implications for Research

Magnification was identified as the link between magic lens type and size. Future research in this area should consider isolating magnification and the dual-view effect. We also observed participants moving close to the AR workspace when merging views using the user-perspective magic lens. This degraded the performance of camera pose tracking, indicating a need for further research into this area.

Implications for Users

We detected that the dual-view problem imposes limitations on the use of surrounding context and identified three attributes that should improve the usability of future handheld AR systems: (1) Optimal magic lens type selection for a given use case and setting; (2) Design device-perspective systems so that they operate magic lenses in regions where the magnification problem is not present; (3) Promote user-perspective view implementations, over device-perspective implementations.

CONCLUSION AND FUTURE WORK

In this paper the effects of the dual-view problem on the use of surrounding context were studied in detail by comparing device-perspective and fixed-POV user-perspective rendering on commercially available handheld devices. Both studies identified a significant effect of the dual-view problem on the use of the surrounding context which was more pronounced for tablet-sized displays. On these large lenses, we pinpointed the differences observed to the cumulative effect of magnification and the dual-view problem. We then derived a set of design guidelines based on the magnification model and minimization graph plot to aid designers in identifying the most appropriate magic lens type for a given task and AR setup. Finally, observational results uncovered different interaction patterns, depending on the view-merging strategy: moving the magic lens closer to AR workspace when trying to simultaneously merge the two views; and further away when using the system as two separate views.

Even with fixed-POV user-perspective rendering, participants struggled to seamlessly merge the two views. To help address this problem, future work should focus on: (1) The diplopia problem, identified as a further factor potentially affecting the use of the surrounding context. (2) Improving camera pose tracking algorithms at close proximity; (3) Improving the accuracy of user-perspective imagery by employing head pose tracking.

ACKNOWLEDGMENTS

Authors would like to thank Slovene Human Resources Development and Scholarship Fund for PhD funding that made carrying this work out possible.

REFERENCES

1. Artur, K.W., Booth, K.S., and Ware, C. Evaluating 3D task performance for fish tank virtual worlds. In *TOIS ‘93*, 239-265.
2. Asteriadis, S., Karpouzis, K., and Kollias, S. Head pose estimation with one camera in uncalibrated environments. In *EGIHMI ’10*, 55-62.
3. Baričević, D., Lee, C., Höllerer, T., and Bowman, D.A. A hand-held AR magic lens with user-perspective rendering. In *ISMAR’12*, 197-206.
4. Baudisch, P., Good, N., Bellotti, V., and Schraedley, P. Keeping things in context : A comparative evaluation of focus plus context screens, overviews, and zooming. In *SIGCHI ‘02*, 259-266.
5. Baudisch, P., Good, N., and Stewart, P. Focus plus context screens: combining display technology with visualization techniques. In *UIST ‘01*, 31-40.
6. Bier, E.A., Stone, M.C., Pier, K., Buxton, W., and DeRose, T. D. Toolglass and magic lenses: the see-through interface. In *SIGGRAPH ’93*, 73-80.
7. Butts, D.R.W and McAllister, D.F. Implementation of true 3D cursors in computer graphics. In proc. *SPIE*, vol. 902, 74-84, 1988.
8. Čopič Pucihar, K., Coulton, P., and Alexander, J. Creating a stereoscopic magic-lens to improve depth perception in handheld augmented reality. In *MobileHCI ’13*, 448-451.
9. Čopič Pucihar, K., Coulton, P., and Alexander, J. Evaluating dual-view perceptual issues in handheld augmented reality: Device vs. user perspective rendering. In *ICMI ’13,* 381-388.
10. Drascic, D. and Milgram, P. Perceptual issues in augmented reality. In proc. *SPIE*, vol. 2653, 123–134.
11. Duwaer, L. and Brink, G. What is the diplopia threshold? Perception & psychophysics, vol. 29, no. 4, 295–309, 1981.
12. Furnas, G. Generalized fisheye views. In *CHI ‘86*, 16-23.
13. Hill, A., Wilson, J., Davidson, B., Gandy, M., and Macintyre, B. Virtual transparency: Introducing parallax view into video see-through AR, In *ISMAR ’11*, 239-240.
14. Jauregui, D.A.G., Argelaguet, F., and Lecuyer, A. Design and evaluation of 3D cursors and motion-parallax for the exploration of desktop virtual environments. In *3DUI ’12*, 69–76.
15. Klein, G. and Murray, D. Parallel tracking and mapping on a camera phone. In *ISMAR ‘09*, 83–86.
16. Kruijff, E., Swan, J.E., and Feiner, S. Perceptual issues in augmented reality revisited. In *ISMAR ’09*, 3-12.
17. Kurz, D. and Benhimane, S. Gravity-aware handheld augmented reality. In *ISMAR ‘11*, 111–120.
18. Liao, C., Liu, Q., Liew, B., and Wilcox, L. Pacer: fine-grained interactive paper via camera-touch hybrid gestures on a cell phone. In *CHI ‘10*, 2441–2450.
19. Matsuda Y., Shibata F., and Tamura A. Creating a user-specific perspective view for mobile mixed reality systems on smartphones. In *3DUI ‘13*, 157–158.
20. Mehra, S., Werkhoven, P., and Worring, M. Navigating on handheld displays: Dynamic versus static peephole navigation. In *TOCHI ‘06*, 448– 457.
21. Oh J. and Hua H. User evaluations on form factors of tangible magic lenses. In *ISMAR ‘06*, 23–32.
22. Olsson T., and Salo M. Online user survey on current mobile augmented reality applications. In *ISMAR ‘11*, 75–84.
23. Robbins, D. C., Cutrell E., Sarin R., and Horvitz E. ZoneZoom: Map navigation for smartphones with recursive view segmentation. In *AVI ‘04*, 231–234.
24. Rohs, M., Schöning, J., Schleicher, R., Essl, G., Naumann, A., and Krüger, A. Impact of Item Density on Magic Lens Interactions. In *MobileHCI ’09*, 1-4.
25. Rohs, M., Schöning, M., Raubal, G., Essl, and Krüger, A. Map navigation with mobile devices: Virtual versus physical movement with and without visual context. In *ICMI’ 07*, 146–153.